EXHIBIT 6

EXHIBIT 6

SENATE JUDICIARY COMMITTEE
Senator Thomas Umberg, Chair
2021-2022 Regular Session

AB 587 (Gabriel)
Version: April 28, 2021
Hearing Date: July 13, 2021
Fiscal: Yes
Urgency: No
CK

SUBJECT

Social media companies: terms of service

DIGEST

This bill requires social media companies, as defined, to post their terms of service and to submit quarterly reports to the Attorney General on their terms of service and content moderation policies and outcomes.

EXECUTIVE SUMMARY

In 2005, five percent of adults in the United States used social media. In just six years, that number jumped to half of all Americans. Today, over 70 percent of adults use at least one social media platform. Facebook alone is used by 69 percent of adults, and 70 percent of those adults say they use the platform on a daily basis.

Given the reach of social media platforms and the role they play in many people's lives, concerns have arisen over what content permeates these sites, entering the lives of the billions of users, and the effects that has on them and society as a whole. In particular, the sharpest calls for action focus on the rampant spread of misinformation, hate speech, and sexually explicit content. Social media companies' content moderation of a decade ago involved handfuls of individuals and user policies were minimal. These programs and policies have dramatically evolved over the years but the proliferation of objectionable content and "fake news" has led to calls for swifter and more aggressive action in response. However, there has also been backlash against perceived censorship in response to filtering of content and alleged "shadow banning."

This bill requires social media companies, as defined, to publicly post their terms of service, with certain required elements, and to provide the Attorney General with a quarterly report on their content moderation procedures and outcomes.

This bill is sponsored by the Anti-Defamation League. It is supported by a variety of groups, including Common Sense and the Islamic Networks Group. It is opposed by various technology and business associations, including the California Chamber of Commerce, the Internet Association, and TechNet.

PROPOSED CHANGES TO THE LAW

Existing law:

1) Prohibits, through the United States Constitution, the enactment of any law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances. (U.S. Const. Amend. 1.)

2) Provides, through the California Constitution, for the right of every person to freely speak, write and publish their sentiments on all subjects, being responsible for the abuse of this right. Existing law further provides that a law may not restrain or abridge liberty of speech or press. (Cal. Const., art. I, § 2(a).)

3) Provides, in federal law, that a provider or user of an interactive computer service shall not be treated as the publisher or speaker of any information provided by another information content provider. (47 U.S.C. § 230(c)(2).)

4) Provides that a provider or user of an interactive computer service shall not be held liable on account of:
    a) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or
    b) any action taken to enable or make available to information content providers or others the technical means to restrict access to such material. (47 U.S.C. § 230(c)(2).)

5) Defines "interactive computer service" as any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions. (47 U.S.C. § 230(f)(2).)

6) Establishes the Unfair Competition Law (UCL) and defines "unfair competition" to mean and include any unlawful, unfair, or fraudulent business act or practice and unfair, deceptive, untrue, or misleading advertising and any act prohibited

by Chapter 1 (commencing with Section 17500) of Part 3 of Division 7 of the Business and Professions Code. (Bus. & Prof. Code § 17200 et seq.)

7) Provides that any person who engages, has engaged, or proposes to engage in unfair competition may be enjoined. Any person may pursue representative claims or relief on behalf of others only if the claimant meets specified standing requirements and complies with Section 382 of the Code of Civil Procedure, but these limitations do not apply to claims brought under this chapter by the Attorney General, or any district attorney, county counsel, city attorney, or city prosecutor in this state. (Bus. & Prof. Code § 17203.)

8) Requires actions for relief pursuant to the UCL be prosecuted exclusively in a court of competent jurisdiction and only by the following:
   a) the Attorney General;
   b) a district attorney;
   c) a county counsel authorized by agreement with the district attorney in actions involving violation of a county ordinance;
   d) a city attorney of a city having a population in excess of 750,000;
   e) a city attorney in a city and county;
   f) a city prosecutor in a city having a full-time city prosecutor in the name of the people of the State of California upon their own complaint or upon the complaint of a board, officer, person, corporation, or association with the consent of the district attorney; or
   g) a person who has suffered injury in fact and has lost money or property as a result of the unfair competition. (Bus. & Prof. Code § 17204.)

9) Holds any person who engages, has engaged, or proposes to engage in unfair competition liable for a civil penalty not to exceed $2,500 for each violation, which shall be assessed and recovered in a civil action brought by the Attorney General, or other public prosecutors. (Bus. & Prof. Code § 17206(a).)

10) Prohibits false or deceptive advertising to consumers about the nature of any property, product, or service, including false or misleading statements made in print, over the internet, or any other advertising method. (Bus. & Prof. Code § 17500.)

11) Defines libel as a false and unprivileged publication by writing, printing, or any other representation that exposes any person to hatred, contempt, ridicule, or obloquy, which causes that person to be shunned or avoided, or which has a tendency to injure that person in their occupation. (Civ. Code §§ 45, 47.)

12) Requires certain businesses to disclose the existence and details of specified policies, including:

a) Operators of commercial websites or online services that collect personally identifiable information about individual consumers residing in California who use or visit the website must conspicuously post its privacy policy. (Bus. & Prof. Code § 22575.)
b) Retailers and manufacturers doing business in this state and having annual worldwide gross receipts over $100,000,000 must disclose online whether the business has a policy to combat human trafficking and, if so, certain details about that policy. (Civ. Code § 1714.43.)
c) End-users of automated license plate recognition technology must post its usage and privacy policy on its website. (Civ. Code § 1798.90.53.)
d) Campus bookstores at public postsecondary educational institutions must post in-store or online a disclosure of its retail pricing policy on new and used textbooks. (Educ. Code § 66406.7(f).)

This bill:

1) Requires a social media company to post their terms of service in a manner reasonably designed to inform all users of the internet-based service owned or operated by the social media company of the existence and contents of the terms of service. The terms of service shall include all of the following:
   a) contact information for the purpose of allowing users to ask the social media company questions about the terms of service;
   b) a description of the process that users must follow to flag content, groups, or other users that they believe violate the terms of service, and the social media company's commitments on response and resolution time; and
   c) a list of potential actions the social media company may take against an item of content or a user, including, but not limited to, removal, demonetization, deprioritization, or banning.

2) Requires the terms of service to be available in all languages in which the social media company offers product features, including, but not limited to, menus and prompts.

3) Provides that a social media company shall be in violation only if the social media company fails to comply within 30 days of being notified of noncompliance by the Attorney General.

4) Requires social media companies to submit a terms of service report, quarterly, with the first report due July 1, 2022, to the Attorney General, who must post it on their website. The terms of service report must include the following:
   a) the current version of the terms of service of the social media company;
   b) if a social media company has filed its first quarterly report, a complete and detailed description of any changes to the terms of service since the last quarterly report;

c) a statement of whether the current version of the terms of service defines specified categories of content, and, if so, the definitions of those categories, including any subcategories. This includes hate speech, racism, extremism, harassment, disinformation, and foreign political interference;
d) a complete and detailed description of content moderation practices used by the social media company, including, but not limited to, all of the following:
    i. any policies intended to address the above categories of content;
    ii. any rules or guidelines regarding how a social media company's automated content moderation systems enforce terms of service and when these systems involve human review;
    iii. any training materials provided to human content moderators intended to educate them on the above categories of content;
    iv. how the social media company responds to user reports of violations of the terms of service;
    v. any rules, guidelines, product changes, and content moderator training materials that cover how the social media company would remove individual pieces of content, users, or groups that violate the terms of service, or take broader action against individual users or against groups of users that violate the terms of service;
    vi. the languages in which the social media company offers product features, and the languages for which the social media company has terms of service;
e) information on content that was flagged by the social media company as content belonging to any of the above categories, including the total number of all of the following:
    i. flagged items of content;
    ii. actioned items of content;
    iii. actioned items of content that resulted in action taken by the social media company against the user or users responsible;
    iv. actioned items of content that were removed, demonetized, or deprioritized by the social media company;
    v. times actioned items of content were viewed by users;
    vi. times actioned items of content were shared, and the number of users that viewed the content before it was actioned; and
    vii. times users appealed social media company actions and the number of reversals on appeal disaggregated by each action;
f) all information required by (e) shall also be disaggregated into the category of content, the type of content, the type of media, and how the content was flagged and actioned.

5) Defines "social media company" as a person or entity that owns or operates a public-facing internet-based service that generated at least $100,000,000 in gross

revenue during the preceding calendar year, and that allows users in the state to do all of the following:

a) construct a public or semipublic profile within a bounded system created by the service;
b) populate a list of other users with whom an individual shares a connection within the system; and
c) view and navigate a list of the individual's connections and the connections made by other individuals within the system.

6) Provides that a "social media company" does not include a person or entity that exclusively owns and operates an electronic mail service.

7) Defines "actioned" to mean a social media company, due to a suspected or confirmed violation of the terms of service, has taken some form of action, including, but not limited to, removal, demonetization, deprioritization, or banning, against the relevant user or relevant item of content.

8) Defines "terms of service" as a policy adopted by a social media company that specifies, at least, the user behavior and activities that are permitted on the internet-based service owned or operated by the social media company, and the user behavior and activities that may subject the user or an item of content to being actioned. This may include, but is not limited to, a terms of service document or agreement, rules or content moderation guidelines, community guidelines, acceptable uses, and other policies and established practices that outline these policies.

9) Makes violations of its provisions actionable under the Unfair Competition Law, Business and Professions Code section 17200 et seq., and any other applicable state or federal law.

COMMENTS

1. Social media content

In recent years, the clamor for more robust content moderation on social media has reached a fever pitch. This includes calls to control disinformation or "fake news," hate speech, political interference, and other online harassment.

The 2016 election was a major breaking point for many. Investigations uncovered attempted interference in the United States Presidential election through a social media "information warfare campaign designed to spread disinformation and societal division

in the United States."[1] The United States Senate Select Committee on Intelligence issued a report detailing how Russian operatives carried out their plan:

> Masquerading as Americans, these operatives used targeted advertisements, intentionally falsified news articles, self-generated content, and social media platform tools to interact with and attempt to deceive tens of millions of social media users in the United States. This campaign sought to polarize Americans on the basis of societal, ideological, and racial differences, provoked real world events, and was part of a foreign government's covert support of Russia's favored candidate in the U.S. presidential election.

This again became a threat in the 2020 election, with social media rife with misinformation such as the incorrect election date,[2] and then social media became a hotbed of misinformation about the results of the election.[3] The author points to investigations that have found the violent insurrectionists that stormed the Capitol on January 6, 2021, were abetted and encouraged by posts on social media sites.[4] In response to indications that social media provided a venue for those who overran and assaulted police officers, Facebook deflected blame, asserting that "these events were largely organized on platforms that don't have our abilities to stop hate, don't have our standards, and don't have our transparency."[5] However, later indictments of those perpetrating the attack "made it clear just how large a part Facebook had played, both in spreading misinformation about election fraud to fuel anger among the Jan. 6 protesters, and in aiding the extremist militia's communication ahead of the riots."[6]

---

[1] Select Committee on Intelligence, Russian Active Measures, Campaigns, and Interference in the 2016 U.S. Election, United States Senate, https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf. All internet citations are current as of July 8, 2021.

[2] Pam Fessler, *Robocalls, Rumors And Emails: Last-Minute Election Disinformation Floods Voters*, NPR (October 24, 2020), https://www.npr.org/2020/10/24/927300432/robocalls-rumors-and-emails-last-minute-election-disinformation-floods-voters.

[3] Sheera Frenkel, *How Misinformation 'Superspreaders' Seed False Election Theories*, New York Times (November 23, 2020), https://www.nytimes.com/2020/11/23/technology/election-misinformation-facebook-twitter.html; Philip Bump, *The chain between Trump's misinformation and violent anger remains unbroken*, Washington Post (May 12, 2021), https://www.washingtonpost.com/politics/2021/05/12/chain-between-trumps-misinformation-violent-anger-remains-unbroken/.

[4] Ken Dilanian & Ben Collins, *There are hundreds of posts about plans to attack the Capitol. Why hasn't this evidence been used in court?* (April 20, 2021) NBC News, https://www.nbcnews.com/politics/justice-department/we-found-hundreds-posts-about-plans-attack-capitol-why-aren-n1264291.

[5] Sheera Frenkel & Cecilia Kang, *Mark Zuckerberg and Sheryl Sandberg's Partnership Did Not Survive Trump* (July 8, 2021) The New York Times, https://www.nytimes.com/2021/07/08/business/mark-zuckerberg-sheryl-sandberg-facebook.html.

[6] *Ibid.*

One area the author specifically focuses in on as motivation for the bill is the rise of hate speech online and the real world consequences. The author points to a recent study of over 500 million Twitter posts from 100 cities in the United States that found that "more targeted, discriminatory tweets posted in a city related to a higher number of hate crimes."[7]

Misinformation also poses a danger to public health: One study found that the more people rely on social media as their main news source, the more likely they are to believe misinformation about the COVID-19 pandemic.[8] Another found that a mere 12 people are responsible for 65 percent of the false and misleading claims about COVID-19 vaccines on Facebook, Instagram, and Twitter.[9] Misinformation hinders emergency responses to natural responses, when social media posts contain incorrect or out-of-date information.[10]

The author frames the problem:

> Over the past several years, there has been growing concern around the role of social media in promoting hate speech, disinformation, conspiracy theories, violent extremism, and severe political polarization. If properly managed, the ability for social media to amplify ideas and messages that would otherwise lack widespread exposure can give voice to otherwise marginalized populations and improve the public discourse, but the same capacity can feed the propagation of misinformation and dangerous rhetoric.

Writing in support, the Anti-Defamation League, the sponsor of this bill, further explains the context of the bill:

> In recent years, there has been growing concern around the role of social media in promoting hate speech, disinformation, conspiracy theories, violent extremism, harassment, and severe political polarization.

---

[7] Press Release, *Hate speech on Twitter predicts frequency of real-life hate crimes* (June 24, 2019) NYU Tanden School of Engineering, https://engineering.nyu.edu/news/hate-speech-twitter-predicts-frequency-real-life-hate-crimes.

[8] Yan Su, *It doesn't take a village to fall for misinformation: Social media use, discussion heterogeneity preference, worry of the virus, faith in scientists, and COVID-19-related misinformation belief* (May 2021) Telematics and Information, Vol. 58, https://www.sciencedirect.com/science/article/abs/pii/S0736585320302069?via%3Dihub.

[9] Shannon Bond, *Just 12 People Are Behind Most Vaccine Hoaxes On Social Media, Research Shows* (May 14, 2021) NPR, https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-twitters-ability-to-curb-vaccine-hoaxes.

[10] United States Department of Homeland Security, *Countering False Information on Social Media in Disasters and Emergencies* (March 2018), https://www.dhs.gov/sites/default/files/publications/SMWG_Countering-False-Info-Social-Media-Disasters-Emergencies_Mar2018-508.pdf.

> According to ADL's 2021 Online Hate and Harassment Survey, 41% of individuals experience online harassment and one in three of those individuals attribute at least some harassment to their identity. Identity-based harassment remains worrisome, affecting the ability of already marginalized communities to be safe in digital spaces.
>
> Importantly, this hate and harassment isn't only taking place in the dark corners of the internet. 75% of ADL's 2021 Online Hate and Harassment Survey respondents who were harassed said at least some harassment happened on Facebook – and many also attributed harassment to other mainstream social media platforms. And online extremism is also front and center: Facebook's own researchers found that 64% of people who joined an extremist group on Facebook only did so because the company's algorithm recommended it to them.

A recent Congressional Research Services Report discussed the issue of content moderation and specifically the spread of misinformation and the role that social media companies play in worsening the issue:

> Two features of social media platforms—the user networks and the algorithmic filtering used to manage content—can contribute to the spread of misinformation. Users can build their own social networks, which affect the content that they see, including the types of misinformation they may be exposed to. Most social media operators use algorithms to sort and prioritize the content placed on their sites. These algorithms are generally built to increase user engagement, such as clicking links or commenting on posts. In particular, social media operators that rely on advertising placed next to user-generated content as their primary source of revenue have incentives to increase user engagement. These operators may be able to increase their revenue by serving more ads to users and potentially charging higher fees to advertisers. Thus, algorithms may amplify certain content, which can include misinformation, if it captures users' attention.[11]

The role that content moderation, or the lack of it, has in alleviating or exacerbating these issues has been a source of much debate. A policy paper published by the Shorenstein Center on Media, Politics, and Public Policy at the Harvard Kennedy School, *Countering Negative Externalities in Digital Platforms*, focuses on the costs associated with various internet platforms that are not absorbed by the companies themselves:

---

[11] Jason A. Gallo & Clare Y. Cho, *Social Media: Misinformation and Content Moderation Issues for Congress* (January 27, 2021) Congressional Research Service, https://crsreports.congress.gov/product/pdf/R/R46662.

> Today, in addition to the carcinogenic effects of chemical runoffs and first and second hand tobacco smoke, we have to contend with a new problem: the poisoning of our democratic system through foreign influence campaigns, intentional dissemination of misinformation, and incitements to violence inadvertently enabled by Facebook, YouTube and our other major digital platform companies.[12]

The paper asserts that these major platform companies "enable exceptionally malign activities" and "experience shows that the companies have not made sufficient investments to eliminate or reduce these negative externalities."

As pointed out by recent Wall Street Journal reporting, the companies' employees are aware of the dangers:

> A Facebook Inc. team had a blunt message for senior executives. The company's algorithms weren't bringing people together. They were driving people apart.
>
> "Our algorithms exploit the human brain's attraction to divisiveness," read a slide from a 2018 presentation. "If left unchecked," it warned, Facebook would feed users "more and more divisive content in an effort to gain user attention & increase time on the platform."
>
> That presentation went to the heart of a question dogging Facebook almost since its founding: Does its platform aggravate polarization and tribal behavior?
>
> The answer it found, in some cases, was yes.[13]

A recent New York Times article on leadership at Facebook elaborates:

> To achieve its record-setting growth, the [Facebook] had continued building on its core technology, making business decisions based on how many hours of the day people spent on Facebook and how many times a day they returned. Facebook's algorithms didn't measure if the magnetic force pulling them back to Facebook was the habit of wishing a friend happy birthday, or a rabbit hole of conspiracies and misinformation.

---

[12] *Countering Negative Externalities in Digital Platforms* (October 7, 2019) Shorenstein Center on Media, Politics and Public Policy, https://shorensteincenter.org/countering-negative-externalities-in-digital-platforms/.

[13] Jeff Horowitz & Deepa Seetharaman, *Facebook Executives Shut Down Efforts to Make the Site Less Divisive* (May 26, 2020) Wall Street Journal, https://www.wsj.com/articles/facebook-knows-it-encourages-division-topexecutives-nixed-solutions-11590507499.

> Facebook's problems were features, not bugs.[14]

Another paper recently released provides "Recommendations to the Biden Administration," and is relevant to the considerations here:

> The Administration should work with Congress to develop a system of financial incentives to encourage greater industry attention to the social costs, or "externalities," imposed by social media platforms. A system of meaningful fines for violating industry standards of conduct regarding harmful content on the internet is one example. In addition, the Administration should promote greater transparency of the placement of digital advertising, the dominant source of social media revenue. This would create an incentive for social media companies to modify their algorithms and practices related to harmful content, which their advertisers generally seek to avoid.[15]

2. Content moderation, transparency, and the low-grade war on our cognitive security

There are a number of considerations when addressing how to approach the proliferation of these undesirable social media posts and the companies' practices that fuel the flames. A number of methods of content moderation are being deployed and have evolved from simply blocking content or banning accounts to quarantining topics, removing posts from search results, barring recommendations, and down ranking posts in priority. However, there is a lack of transparency and understanding of exactly what companies are doing and why it does not seem to be enough. A recent article in the MIT Technology Review articulates the issues with content moderation behind the curtain:

> As social media companies suspended accounts and labeled and deleted posts, many researchers, civil society organizations, and journalists scrambled to understand their decisions. The lack of transparency about those decisions and processes means that—for many—the election results end up with an asterisk this year, just as they did in 2016.
>
> What actions did these companies take? How do their moderation teams work? What is the process for making decisions? Over the last few years,

---

[14] Sheera Frenkel & Cecilia Kang, *Mark Zuckerberg and Sheryl Sandberg's Partnership Did Not Survive Trump* (July 8, 2021) The New York Times, https://www.nytimes.com/2021/07/08/business/mark-zuckerberg-sheryl-sandberg-facebook.html.

[15] Caroline Atkinson, et al., *Recommendations to the Biden Administration On Regulating Disinformation and Other Harmful Content on Social Media* (March 2021) Harvard Kennedy School & New York University Stern School of Business, https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/6058a456ca24454a73370dc8/1616421974691/TechnologyRecommendations_2021final.pdf.

> platform companies put together large task forces dedicated to removing election misinformation and labeling early declarations of victory. Sarah Roberts, a professor at UCLA, has written about the invisible labor of platform content moderators as a shadow industry, a labyrinth of contractors and complex rules which the public knows little about. Why don't we know more?
>
> In the post-election fog, social media has become the terrain for a low-grade war on our cognitive security, with misinformation campaigns and conspiracy theories proliferating. When the broadcast news business served the role of information gatekeeper, it was saddled with public interest obligations such as sharing timely, local, and relevant information. Social media companies have inherited a similar position in society, but they have not taken on those same responsibilities. This situation has loaded the cannons for claims of bias and censorship in how they moderated election-related content.

This bill seeks to increase transparency around what terms of service social media companies are setting out and how it ensures those terms are abided by. The goal is to learn more about the methods of content moderation and how successful they are. According to the author:

> The line between providing an open forum for productive discourse and permitting the proliferation of hate speech and misinformation is a fine one, and depends largely on the structure and practices of the platform. However, these platforms rarely provide detailed insight into such practices, and into the relative effectiveness of different approaches. This, along with constraints imposed by existing federal law, has historically made policy-making in this space remarkably difficult. This bill seeks to provide critical transparency to both inform the public as to the policies and practices governing the content they post and engage with on social media, and to allow for comparative assessment of content moderation approaches to better equip both social media companies and policymakers to address these growing concerns.

ADL emphasizes the need for the bill:

> Despite the widespread nature of these concerns, efforts by social media companies to self-police such content have been opaque, arbitrary, biased, and inadequate. While some platforms share limited information about their efforts, the current lack of transparency has exacerbated concerns about the intent, enforcement, and impact of corporate policies. Consequently, policymakers and the general public remain deprived of critical data and metrics regarding the scope and scale of online hate and

> disinformation. Additional transparency is needed to allow consumers to make informed choices about the impact of these products (including on their children) and so that researchers, civil society leaders, and policymakers can take meaningful action to decrease online hate and extremism, and to address this growing threat to our democracy.
>
> The creation of a thoughtful and standardized enumeration and measurement of policies and enforcement will serve policymakers and the public. We need this critical information to better understand the policies and practices of social media platforms – which have a profound impact on communication and discourse.
>
> AB 587 will address this troubling lack of transparency by requiring social media platforms to publicly disclose their corporate polices and report key data and metrics around the enforcement of their policies. This disclosure would be accomplished through regular public filings with the Attorney General.

This bill starts with a baseline requirement to have social media companies post their terms of service. These policies must include information about how users can ask questions, how they can flag content or users in violation, and a list of potential actions that the company might take in response. To ensure meaningful access, the terms of service must be posted in a manner reasonably designed to inform all users of their existence and contents and available in all languages in which the company offers product features. The Attorney General must provide a 30-day right to cure before taking action against companies for failing to abide by these requirements.

The bill next requires an extremely detailed report to be compiled by these companies and submitted to the Attorney General on a quarterly basis. This report must include information on the terms of service, any changes made and whether they define certain categories of content, including hate speech or racism; extremism or radicalization; disinformation or misinformation; harassment; and foreign political interference.

The bill also requires the report to contain a "complete and detailed description of content moderation practices" used by the company. There must also be outcome-focused information included. Companies must report on the number of flagged items of content and the number of times the company took action in response. To understand the impact of the reported content, the report must detail the number of times this content was viewed and shared by users. The data must also include these details broken down by content category, the type of media, and other factors.

As the author references above, all of this occurs within tight quarters due to federal statutory and constitutional law. Section 230 of the Communications Decency Act, in relevant part, immunizes providers from liability for actions taken in good faith "to

restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected." (47 U.S.C. § 230.) It also makes clear that platforms cannot "be treated as the publisher or speaker of any information provided by another information content provider." Section 230's language is intended to provide a broad immunity to incentivize activity that moderates such objectionable content but does not hold companies liable for what is left posted. State laws that are inconsistent with the scheme set out by Section 230 are expressly preempted.

The author argues that the bill walks this line carefully:

> AB 587 does not impose liability based on the nature of content moderation decisions taken by social media platforms. Rather, the requirements of AB 587 are focused exclusively on disclosure of information relating to those practices, with liability imposed based on failure to disclose the specified information. By taking this transparency approach, AB 587 is thus unlikely to run afoul of the liability protections provided by Section 230, and would be far less susceptible to a preemption challenge than most attempts to regulate in this space.

In addition, any specific mandates to remove some subset of this broad swath of content could run afoul of the First Amendment. The United States Supreme Court has held that posting on social networking and/or social media sites constitutes communicative activity protected by the First Amendment.[16] As a general rule, the government "may not suppress lawful speech as the means to suppress unlawful speech."[17] In addition, the First Amendment places restrictions on compelled speech. However, the case law general affords a wide berth to laws that regulate commercial speech and that involve disclosure requirements that involve conveying factual information that has sound public policy justification, such as food labeling.[18]

Because this bill simply seeks transparency into what content moderation practices are being deployed and their outcomes, it likely does not run afoul of these laws. No specific content moderation is required or penalized. The information required to be disclosed can play a key role in informing future legislative action and public debate of these issues.

---

[16] *E.g.*, *Packingham v. North Carolina* (2017) 137 S.Ct. 1730, 1735-1736.

[17] *Ashcroft v. Free Speech Coalition* (2002) 535 U.S. 234, 255; see also *United States v. Alvarez* (2012) 567 U.S. 709, 717 (Supreme Court "has rejected as 'startling and dangerous' a 'free-floating test for First Amendment coverage…[based on] an ad hoc balancing of relative social costs and benefits' " [alterations in original]).

[18] *See Central Hudson Gas & Elec. Corp. v. Public Serv. Comm'n* (1980) 447 U.S. 557; *Zauderer v. Office of Disciplinary Counsel of Supreme Court* (1985) 471 U.S. 626.

3. Defining social media

The bill defines "social media company" as a person or entity that owns or operates a public-facing internet-based service that generated at least $100 million in gross revenue during the preceding calendar year, and that allows users in the state to do all of the following:

- construct a public or semipublic profile within a bounded system created by the service;
- populate a list of other users with whom an individual shares a connection within the system; and
- view and navigate a list of the individual's connections and the connections made by other individuals within the system.

The bill specifically excludes from this definition persons or entities that exclusively own and operate an electronic mail service.

A number of companies and their representative organizations have argued that the definition unnecessarily sweeps up companies that most people would not consider "social media." They request specific exemptions to carve them out of the definition in the bill.

There is evidence that an easy definition is elusive for many. Federal law does not provide a clear definition. For instance, the federal statute establishing the Social Media Working Group within the Department of Homeland Security does not even provide a definition. (6 USC 195d.) Other states that define social media in statute have used definitions similar to that laid out in the bill.[19]

In response to stakeholder concerns and at the request of the Committee to harmonize the relevant definitions in this bill and in AB 35 (Chau, 2021), the author has proposed a revised definition for social media company. One concern raised was that, while a company may operate a social media platform as part of its company, the entire company should not necessarily be subject to the provisions of this bill. The amendments make clear that the requirements of the bill apply to "social media platforms" which are owned or operated by a "social media company." The definition stands apart from a clarifying section that follows, which simply indicates the entities that are not subject to the provisions of this particular bill. The proposed amendments are included at the end of this analysis.

[19] See e.g., N.M. Stat. Ann. § 21-1-46; La. Rev. Stat. Ann. § 3:2462.

4. Concerns with the bill

Three major areas of concern that have been raised, primarily by a coalition in opposition to the bill, are the granularity of the data required to be reported, the reporting scheme itself, and the enforcement provisions. In response to the varied concerns of stakeholders, the author has proposed a series of amendments to the bill, which are included at the end of this analysis.

a. Narrowing what is included in the quarterly reports

The technology and business coalition in opposition, including TechNet and the California Chamber of Commerce, argue:

> In seeking to increase transparency around content moderation practices, AB 587 requires companies to report to the Attorney General the guidelines, practices, and even training materials companies use to moderate their platforms. The recent amendments make it explicit that the bill is seeking "complete and detailed" information about content moderation practices, capabilities, and data regarding content moderation. This requirement would not only threaten the security of these practices but provides bad actors with roadmaps to get around our protections. We believe that while well intentioned, these requirements will ultimately allow scammers, spammers, and other bad actors to exploit our systems and moderators.
>
> To avoid undermining the goals of the bill and the work our companies have already undertaken to combat harmful content, we suggest removing these requirements in order to prevent the disclosure of information that could be used against our platforms and our users.

Although opposition points out that many social media companies already post similar reports on their own websites, supporters believe more is necessary. Decode Democracy writes in support of requiring more:

> While some platforms share limited information about their efforts, more transparency is needed to address concerns about the intent, enforcement, and impact of platform policies, and to provide policymakers and the general public with critical data and metrics regarding the scope and scale of online hate and disinformation. Greater transparency is needed so consumers can make informed choices about the impact of platform policies, (including on their children), and also to enable researchers, civil society leaders, and policymakers to determine the best way to address this threat to our democracy.

Ultimately, the goal of transparency in content moderation and terms of service is to understand what is being done in order to assess its effectiveness and to guide debate about what can be done to enhance that effectiveness. Arguably, this goal is undermined if every tactic to combat objectionable content, such as misinformation, hate speech, and outright criminal activity, was divulged in complete detail to those intending to subvert those very practices. Although the sophisticated actors that carry a majority of this out will likely already have much of this information, a better balance can be struck in the bill. In response, the author proposes a number of amendments to scale back the granularity of the data required to be placed in these reports. As seen by the language included at the end of this analysis, the author proposes removing the requirement to include items such as training manuals, a complete description of the company's rules and guidelines for content moderation, among other data points.

b. Reporting requirements

The coalition in opposition next argues the reporting requirements are too burdensome:

> AB 587 requires businesses to report detailed metrics on a quarterly basis regarding not only the numerical scale of content moderation practices, but also details about how content is flagged and acted against. It would be nearly impossible to report this information quarterly due to the need to review, analyze, and adjudicate actioned content. Further, the sheer volume of content our companies review makes it similarly difficult and costly to implement these disclosures, particularly the number of times actioned items of content were viewed or shared. Producing this information quarterly is unworkable and unreasonable.
>
> There is little justification to require a report to the Attorney General. Instead the bill should simply require our companies to post these reports on their website or platform.

The amendments discussed in the previous section will certainly mitigate the burden these reports impose. However, it is arguably unnecessarily onerous to require all of this information to be reported every three months. The author may wish to consider moving to longer reporting periods.

c. Enforcement

The bill provides that a violation of any of its provisions is actionable under the Unfair Competition Law. The requirement that a social media company post their terms of service, and include certain details, comes with a right to cure. Therefore, a company is not in violation of that section unless it fails to comply within 30 days of the Attorney General notifying them of their noncompliance.

The coalition in opposition asserts that the enforcement mechanisms are onerous and problematic, arguing against even allowing for injunctive relief. The coalition states:

> AB 587 opens companies up to the threat of liability and government investigation for routine moderation practices. Companies should not be subject to civil penalties or injunctive relief for the filing of a report, especially as comprehensive as the ones contemplated by this bill. Such litigation will deter investment in content moderation and suppress ongoing efforts to protect users from harmful content online.

The bill simply seeks more transparency for better-informed decisionmaking on appropriate next steps. However, this crucial information will likely be hard to gather if there is no meaningful penalty for failing to comply. Despite the opposition's arguments, the provision providing for enforcement pursuant to the UCL is likely not even necessary, as violation of the bill's provisions would serve as a predicate offense for a UCL action. It should be noted that it is improbable that any private right of action would be afforded by this provision, given the nature of potential injuries. Therefore, enforcement is left to public prosecutors and counsel for local governments.

In addition, the right to cure before the Attorney General can enforce the requirement regarding posting terms of service is arguably unwarranted. Allowing noncompliance with the law until 30 days after the Attorney General investigates and determines there is noncompliance is not sound public policy and undermines meaningful enforcement. Given the repeated assertions by those in opposition that social media companies "already make their terms of service and community standards easily accessible on their websites," it does not seem an overly onerous requirement and noncompliance should be sufficient for the Attorney General to enforce within its discretion.

Ultimately, to ensure clarity in how enforcement is to be carried out, the author has proposed amendments, included below, that remove the right to cure and the UCL provisions. The proposed amendments insert a penalty scheme involving civil penalties for each violation for every day the violation continues.

SUPPORT

Anti-Defamation League (sponsor)
American Association of University Women, Camarillo Branch
Accountable Tech
American Jewish Committee - Los Angeles
The Arc and United Cerebral Palsy California Collaboration
Armenian Assembly of America
Armenian National Committee of America - Western Region
Asian Americans in Action
Bend the Arc: Jewish Action

Buen Vecino
California Asian Pacific American Bar Association
California Hawaii State Conference National Association for the Advancement of Colored People
California League of United Latin American Citizens
Center for the Study of Hate & Extremism - California State University, San Bernardino
Common Sense
Decode Democracy
Esperanza Immigrant Rights Project, Catholic Charities of Los Angeles
The Greenlining Institute
Hindu American Foundation, Inc.
Islamic Networks Group
Israeli-American Civic Action Network
Japanese American Citizens League, Berkeley Chapter
Jewish Center for Justice
Jewish Public Affairs Committee
Korean American Bar Association of Northern California
Korean American Coalition - Los Angeles
League of United Latin American Citizens
National Association for the Advancement of Colored People, SV/SJ
Nailing It for America
National Council of Jewish Women, California
National Hispanic Media Coalition
Orange County Racial Justice Collaborative
Progressive Zionists of California
Rabbis and Cantor of Congregation or Ami
Sikh American Legal Defense and Education Fund (SALDEF)
Simon Wiesenthal Center, Inc.
Stonewall Democratic Club

OPPOSITION

California Chamber of Commerce
Chamber of Progress
Civil Justice Association of California
Consumer Technology Association
Internet Association
Internet Coalition
MPA - the Association of Magazine Media
Netchoice
TechNet

RELATED LEGISLATION

Pending Legislation:

SB 388 (Stern, 2021) requires a social media platform company, as defined, that, in combination with each subsidiary and affiliate of the service, has 25,000,000 or more unique monthly visitors or users for a majority of the preceding 12 months, to report to the Department of Justice by April 1, 2022, and annually thereafter, certain information relating to its efforts to prevent, mitigate the effects of, and remove potentially harmful content. SB 388 is pending before the Senate Judiciary Committee.

SB 435 (Cortese, 2021) provides, in relevant part, for a cause of action against an entity that publishes or republishes certain sexual content, as provided, and provides for civil penalties for every two hours flagged content is not taken down, as specified. SB 435 is pending before the Senate Judiciary Committee.

SB 746 (Skinner, 2021) requires businesses to disclose whether they use the personal information of consumers for political purposes, as defined, to consumers, upon request, and annually to the Attorney General or the California Privacy Protection Agency. This bill is currently on the Senate Floor.

AB 35 (Chau, 2021) requires a person that operates a social media platform to disclose whether or not the platform has a policy or mechanism in place to address the spread of misinformation. AB 35 is currently in this Committee and will be heard on the same day as this bill.

AB 1379 (Eduardo Garcia, 2021) requires an online platform that has 10,000,000 or more unique monthly United States visitors or users for a majority of months during the preceding 12 months that targets political advertising, as defined, to make available an application programming interface or other technical capability to enable qualified third parties to conduct independent analysis of bias and unlawful discriminatory impact of that targeted advertising. AB 1379 is pending before the Assembly Elections Committee.

AB 1114 (Gallagher, 2021) requires a social media company located in California to develop a policy or mechanism to address content or communications that constitute unprotected speech, including obscenity, incitement of imminent lawless action, and true threats, or that purport to state factual information that is demonstrably false. AB 1114 is pending before the Assembly Arts, Entertainment, Sports, Tourism, and Internet Media Committee.

AB 613 (Cristina Garcia, 2021) requires social media platforms, as defined, or users or advertisers posting on a social media platform, to place text or marking within or adjacent to retouched images that have been posted on the platform for promotional or

commercial purposes, and specify how that retouched image was altered. AB 613 is pending before the Assembly Privacy and Consumer Protection Committee.

Prior Legislation:

SB 890 (Pan, 2020) would have required social media companies to remove images and videos depicting crimes, as specified, and imposed civil penalties for failing to do so. SB 890 died in the Senate Judiciary Committee.

AB 2391 (Gallagher, 2020) would have prohibited social media sites from removing user-posted content on the basis of the political affiliation or viewpoint of that content, except where the social media site is, by its terms and conditions, limited to the promotion of only certain viewpoints and values and the removed content conflicts with those viewpoints or values. AB 2931 died in the Assembly Committee on Arts, Entertainment, Sports, Tourism, and Media.

AB 2442 (Chau, 2020) was substantially similar to this bill and would have required social media companies to disclose the existence, or lack thereof, of a misinformation policy, and imposed civil penalties for failing to do so. AB 2442 died in the Senate Judiciary Committee due to the COVID-19 pandemic.

AB 1316 (Gallagher, 2019) would have prohibited social media sites from removing user-posted content on the basis of the political affiliation or viewpoint of that content, except where the social media site is, by its terms and conditions, limited to the promotion of only certain viewpoints and values and the removed content conflicts with those viewpoints or values. AB 1316 was held on the floor of the Assembly and was re-introduced as AB 2931 (2020).

AB 288 (Cunningham, 2019) would have required a social networking service, at the request of a user, to permanently remove personally identifiable information and not sell the information to third parties, within a commercially reasonable time of the request. AB 288 died in the Assembly Committee on Privacy and Consumer Protection.

SB 1424 (Pan, 2018) would have established a privately funded advisory group to study the problem of the spread of false information through Internet-based social media platforms, and draft a model strategic plan for Internet-based social media platforms to use to mitigate this problem. SB 1424 was vetoed by Governor Brown, whose veto message stated that, as evidenced by the numerous studies by academic and policy groups on the spread of false information, the creation of a statutory advisory group to examine this issue is not necessary.

AB 3169 (Gallagher, 2018) would have prohibited social media sites from removing content on the basis of the political affiliation or viewpoint of the content, and prohibited internet search engines from removing or manipulating content from search

results on the basis of the political affiliation or viewpoint of the content. AB 3169 died in the Assembly Committee on Privacy and Protection.

SB 1361 (Corbett, 2010) would have prohibited social networking websites from displaying, to the public or other registered users, the home address or telephone number of a registered user of that site who is under 18 years of age, and imposed a civil penalty of up to $10,000 for each willful and knowing violation of this prohibition. SB 1361 died in the Assembly Committee on Entertainment, Sports, Tourism, and Internet Media.

PRIOR VOTES:

Assembly Floor (Ayes 64, Noes 1)
Assembly Appropriations Committee (Ayes 13, Noes 0)
Assembly Judiciary Committee (Ayes 10, Noes 0)
Assembly Privacy and Consumer Protection Committee (Ayes 9, Noes 0)

****

AUTHOR'S PROPOSED AMENDMENTS

SEC. 2. Chapter 22.8 (commencing with Section 22675) is added to Division 8 of the Business and Professions Code, to read:

CHAPTER 22.8. Content Moderation Requirements for Internet Terms of Service

22675. For purposes of this chapter, the following definitions apply:

(a) "Actioned" means a social media company, due to a suspected or confirmed violation of the terms of service, has taken some form of action, including, but not limited to, removal, demonetization, deprioritization, or banning, against the relevant user or relevant item of content.

(b) "Content" means media, including, but not limited to, text, images, videos, and groups of users that are created, posted, shared, or otherwise interacted with by users on an internet-based service.

(c) ~~(1)~~ "Social media company" means a person or entity that owns or operates ~~a~~ **one or more social media platforms.**

(d) (1) "Social media platform" means an ~~*public facing*~~ internet-based service that ~~*generated at least one hundred million dollars ($100,000,000) in gross revenue during the preceding calendar year, and that*~~ allows users ~~*in the state*~~ to do all of the following:

(A) Construct a public or semipublic profile within a bounded system created by the service.

(B) Populate a list of other users with whom an individual shares a connection within the system.

(C) View and navigate a list of ~~*the individual's connections and the*~~ connections made by other individuals within the system.

~~*(2) "Social media company" does not include a person or entity that exclusively owns and operates an electronic mail service.*~~

(d) "Terms of service" means a policy **or set of policies** adopted by a social media company that specifies, at least, the user behavior and activities that are permitted on the internet-based service owned or operated by the social media company, and the user behavior and activities that may subject the user or an item of content to being actioned. This may include, but is not limited to, a terms of service document or

agreement, rules or content moderation guidelines, community guidelines, acceptable uses, and other policies and established practices that outline these policies.

22676. (a) A social media company shall post ~~*their*~~ terms of service **for each social media platform owned or operated by that company** in a manner reasonably designed to inform all users of the **social media platform**~~*internet-based service owned or operated by the social media company*~~ of the existence and contents of the terms of service.

(b) The terms of service posted pursuant to subdivision (a) shall include all of the following:

(1) Contact information for the purpose of allowing users to ask the social media company questions about the terms of service.

(2) A description of the process that users must follow to flag content, groups, or other users that they believe violate the terms of service, and the social media company's commitments on response and resolution time.

(3) A list of potential actions the social media company may take against an item of content or a user, including, but not limited to, removal, demonetization, deprioritization, or banning.

(c) The terms of service posted pursuant to subdivision (a) shall be available in all **Medi-Cal threshold** languages**, as defined in subdivision (d) of Section 128552 of the Health and Safety Code,** in which the social media **platform**~~*company*~~ offers product features, including, but not limited to, menus and prompts.

~~*(d) A social media company shall be in violation of this section if the social media company fails to comply with the provisions of this section within 30 days of being notified of noncompliance by the Attorney General.*~~

22677. (a) On a quarterly basis, a social media company shall submit to the Attorney General a terms of service report, covering activity within the three months previous to the submission of the report. The terms of service report shall include**, for each social media platform owned or operated by the company, all of** the following:

(1) The current version of the terms of service of the social media **platform**~~*company*~~.

(2) If a social media company has filed its first quarterly report, a complete and detailed description of any changes to the terms of service since the last quarterly report.

(3) A statement of whether the current version of the terms of service defines each of the following categories of content, and, if so, the definitions of those categories, including any subcategories:

(A) Hate speech or racism.

(B) Extremism or radicalization.

(C) Disinformation or misinformation.

(D) Harassment.

(E) Foreign political interference.

(4) A ~~*complete and*~~ detailed description of content moderation practices used by the social media company **for that platform**, including, but not limited to, all of the following:

(A) Any existing policies intended to address the categories of content described in paragraph (3).

(B) ~~*Any rules or guidelines regarding how*~~ **How** ~~*a social media company's*~~ automated content moderation systems enforce terms of service **of the social media platform** and when these systems involve human review.

~~*(C) Any training materials provided to human content moderators intended to educate them on the categories of content described in paragraph (3).*~~

(D) How the social media company responds to user reports of violations of the terms of service.

(E) ~~*Any rules, guidelines, product changes, and content moderator training materials that cover how*~~ **How** the social media company would remove individual pieces of content, users, or groups that violate the terms of service, or take broader action against individual users or against groups of users that violate the terms of service.

(F) The languages in which the social media ~~*company*~~ **platform does not make terms of service available, but does offer**~~*offers*~~ product features, including, but not limited to, menus and prompts~~*and the languages for which the social media company has terms of service*~~.

(5) (A) Information on content that was flagged by the social media company as content belonging to any of the categories described in paragraph (3), including all of the following:

(i) The total number of flagged items of content.

(ii) The total number of actioned items of content.

(iii) The total number of actioned items of content that resulted in action taken by the social media company against the user or group of users responsible for the content.

(iv) The total number of actioned items of content that were removed, demonetized, or deprioritized by the social media company.

(v) The number of times actioned items of content were viewed by users.

(vi) The number of times actioned items of content were shared, and the number of users that viewed the content before it was actioned.

(vii) The number of times users appealed social media company actions **taken on that platform** and the number of reversals of social media company actions on appeal disaggregated by each type of action.

(B) All information required by subparagraph (A) shall be disaggregated into the following categories:

(i) The category of content, including any relevant categories described in paragraph (3).

(ii) The type of content, including, but not limited to, posts, comments, messages, profiles of users, or groups of users.

(iii) The type of media of the content, including, but not limited to, text, images, and videos.

(iv) How the content was flagged, including, but not limited to, flagged by company employees or contractors, flagged by artificial intelligence software, flagged by community moderators, flagged by civil society partners, and flagged by users.

(v) How the content was actioned, including, but not limited to, actioned by company employees or contractors, actioned by artificial intelligence software, actioned by community moderators, actioned by civil society partners, and actioned by users.

(b) A social media company shall submit its first terms of service report pursuant to subdivision (a) to the Attorney General no later than July 1, 2022.

(c) The Attorney General shall post on its official website all terms of service reports submitted pursuant to this section.

22678. (a) A social media company that violates the provisions of this chapter shall be liable for a civil penalty not to exceed fifteen thousand dollars ($15,000) per violation per day, and may be enjoined in any court of competent jurisdiction.

(b) Actions for relief pursuant to this chapter shall be prosecuted exclusively in a court of competent jurisdiction by the Attorney General or a district attorney or by a county counsel authorized by agreement with the district attorney in actions involving violation of a county ordinance, or by a city attorney of a city having a population in excess of 750,000, or by a city attorney in a city and county or, with the consent of the district attorney, by a city prosecutor in a city having a full-time city prosecutor in the name of the people of the State of California upon their own complaint or upon the complaint of a board, officer, person, corporation, or association.

(c) If an action pursuant to this section is brought by the Attorney General, one-half of the penalty collected shall be paid to the treasurer of the county in which the judgment was entered, and one-half to the General Fund. If the action is brought by a district attorney or county counsel, the penalty collected shall be paid to the treasurer of the county in which the judgment was entered. If the action is brought by a city attorney or city prosecutor, one-half of the penalty collected shall be paid to the treasurer of the city in which the judgment was entered, and one-half to the treasurer of the county in which the judgment was entered.

22679. (a) The duties and obligations imposed by this chapter are cumulative to any other duties or obligations imposed under local, state, or federal law and shall not be construed to relieve any party from any duties or obligations imposed under law.

(b) The remedies or penalties provided by this chapter are cumulative to each other and to any other remedies or penalties available under local, state, or federal law.

~~*22678. A violation of this chapter is actionable under the Unfair Competition Law (Chapter 5 (commencing with Section 17200) of Part 2 of Division 7), in addition to any other applicable state or federal law.*~~

22680. This chapter shall not apply to any of the following:

(a) A social media company that generated less than one hundred million dollars ($100,000,000) in gross revenue during the preceding calendar year.

(b) A service that exclusively conveys electronic mail.

(c) A service that exclusively facilitates direct messaging between users.

(d) A section for user-generated comments on a digital news internet website that otherwise exclusively hosts content published by a person or entity described in subdivision (b) of Section 2 of Article I of the California Constitution.

(e) Consumer reviews of products or services on an internet website that serves the exclusive purpose of facilitating online commerce.

(f) An internet-based subscription streaming service that is offered to consumers for the exclusive purpose of transmitting licensed media, including audio or video files, in a continuous flow from the internet-based service to the end user, and does not host user-generated content.

(g) A service that operates for the exclusive purpose of cloud storage or shared document or file collaboration.

-0-